

## BIBLIOTHEKEN

Webdoc ist ein Service von Bibliotheken, der mittels zentraler Erschließung und zentral implementierter Lizenzierungsmechanismen den Endnutterzugriff auf verteilte elektronische Dokumentressourcen erlaubt. Webdoc ist aber auch eine Chance zur Zusammenarbeit von Verlagen und Bibliotheken bei der digitalen Literaturbereitstellung.

Dr. Stefan Gradmann,  
Direktor der  
Verbundzentrale des GBV,  
über die Hintergründe  
dieses Projekts

scape oder Mosaic, diese sind zudem entweder sehr preiswert oder sogar frei verfügbar.

•// Drittens – und dies ist wohl der entscheidende Faktor – materialisiert das WWW mit seinem Hypertext-Ansatz eine omnipräsente und äußerst griffige Technikmetapher: das Bild der Welt als Text.

In dieser Technikmetapher des weltumspannenden Textes ist der Grundgedanke einer universellen virtuellen Kopräsens und beliebigen Verbindbarkeit aller Einzelelemente bestimmend. Konstitutiv und dominant sind dabei syntagmatische, assoziative Relationen.

Stark zurückgedrängt und (jedenfalls in der noch andauernden Euphoriephase) von untergeordneter Be-

fällig strukturierter Meta-Informationsschichten.

Meta-Informationen in Gestalt von Katalogen, Inventarlisten und finding lists aller Art sind dem durch flache Verknüpfungsbeziehungen bestimmten Textparadigma des WWW in seiner reinen Form im Grunde fremd. Nachträglich unternommene Versuche, dem WWW solche Metastrukturen einzuziehen, haben zumindest lange Zeit immer halbherzigen und unfertigen Charakter gehabt.

Die Erfolgsgeschichte des WWW hat jedoch auch ihre Schattenseite. Jeder Nutzer des Web dürfte Erfahrungen mit dessen Schwächen und Problemen gemacht haben, die mit dem zunehmenden Wachstum dieses weltumspannenden Netzes immer deutlicher sichtbar werden und sich zu ech-

# Reaktion auf

Das Internet ist sicher einer der publizistischen Shooting-Stars der letzten Jahre. Eine erstaunliche Tatsache: Handelt es sich doch im Kern um kaum mehr als ein Bündel integrierender Kommunikationsprotokolle, trockene und dürre Materie und als solche kaum angetan, breite Massen zu faszinieren.

Eine der Hauptursachen für den in den Medien tobenden Internet-Hype ist wohl der für die Gestaltung von Informationsangeboten im Umfeld des Internet derzeit dominante Ansatz: das World Wide Web (WWW). Im World Wide Web gewinnt das ansonsten nur diffus erahnte Potential des weltumspannenden Meta-Netzes plakativ Gestalt. Für das Verständnis der Überlegungen, die dem Webdoc-Projekt zugrunde liegen, ist es entscheidend, sich die Schlüsselfaktoren für den Erfolg des WWW und der Web-orientierten Techniken klarzumachen. Der WWW-Ansatz verdankt seinen Erfolg drei Komponenten:

•// Dienstangebote im WWW basieren auf weit verbreiteten und akzeptierten de-facto-Standards: HTTP als integrierendes anwendungsorientiertes Protokoll und HTML als korrespondierendes Dokumentformat.

•// Dienste im WWW sind nutzbar über generische Clientsoftware in Gestalt der Web-Browser wie Net-

scape oder Mosaic, diese sind zudem entweder sehr preiswert oder sogar frei verfügbar.

Die eher blasse Rolle solcher Konzepte mag zugleich einige der spezifischen Schwierigkeiten erklären, die Bibliotheken bei der Nutzung dieses neuen Informationsmodells haben. Denn gerade solche stärker paradigmatisch orientierten klassischen Modelle und Methoden der Informationsstrukturierung sind von alters her tragende Säulen bibliothekarischer Arbeit gewesen:

•// Linearisierung und Sequentialisierung von Informationseinheiten im Denkmodell des Katalogs

•// Hierarchische Strukturierung im Sinne eines systematisierenden Erschließungsansatzes

•// Statische Lokalisierung im Sinne des Bestandsnachweises.

Solche klassischen Stärken der Bibliotheken als Anbieter und Vermittler von Informationen also kommen im »flachen« Informationsansatz des Web erst einmal kaum zum Tragen. Gleiches gilt für ein weiteres, im Bibliotheksbereich omnipräsentes Konzept: die tragende Rolle viel-

ten Einschränkungen entwickeln. Wer hat nicht schon auf der Suche nach einer ganz bestimmten Information Stunden im Web zugebracht und stand am Ende einer solchen Session mit einer ganzen Fülle bunter, je für sich genommen sicher faszinierender Informationsschnipsel da – nur daß die ursprünglich gesuchte Information nicht darunter war (wohl dem, der sich dann nach Stunden wenigstens noch daran erinnert, was er denn anfangs gesucht hatte). Oder – eine traumatische Grunderfahrung – ein am Vortag auf einem Serversystem georteter Text ist plötzlich etwas zu virtuell geworden, und es gibt ihn nur noch als Bookmark im eigenen Browser. Entweder ist das ursprünglich referenzierte Objekt schlicht verschwunden (wenigstens sind dann die Folgen noch klar zu bestimmen: Die Information ist schlicht verloren). Oder – schlimmer noch – das referenzierte Dokument scheint irgendwie verändert, ohne daß sich präzise sagen ließe, wann und in welchem Umfang diese Veränderungen vorgenommen wurden. Ob der Link tatsächlich noch auf dasselbe Objekt verweist, wird dann eher zur metaphysischen Frage.

Schließlich die verwirrende Vielfalt im Bereich der laufend zunehmenden kommerziellen Angebote auch im Bereich der Informationsdienste: Hier ist eine Kreditkarteninformation ge-



fordert, an anderer Stelle ist cash gewünscht (meist in US-Dollar und mitunter ohne korrekte Rechnungsstellung: Horrorvision für haushaltsführende Stellen). Und das schlimmste: Bei alledem bleibt (aus Nutzersicht) allzuhäufig unklar, was man schließlich für sein Geld erhalten wird beziehungsweise ob man überhaupt einen realen Gegenwert erhält.

Das Webdoc-Projekt ist eine Reaktion auf diese drei Problemfelder im Bereich der WWW-basierten Dienste:

:// fehlende oder schwach ausgeprägte Dienste für die Strukturierung des Informationsangebots und fehlende Metainformation

:// ungesicherte Verfügbarkeit dieses Informationsangebots beziehungsweise fehlende Authentifizierungsmechanismen

Das technisch-funktionale Konzept von Webdoc (siehe hierzu Funktionsdiagramm) integriert über eine gemeinsame Nachweisdatenbank (Webcat) den Zugriff auf räumlich verteilt angebotene elektronische Dokumente. Zum Einsatz kommen dabei generische und Pica-eigene Komponenten.

Die am Projekt beteiligten Bibliotheken, Service-Provider und Verlage bauen dabei lokale Web-Server auf. Diese Server halten Dokumente in derzeit fünf für die Projektlaufzeit vereinbarten Standardformaten vor. Unterstützt werden dabei die Formate HTML, Postscript, TIFF (mit Kompression gemäß Fax Gruppe 4), PDF und TEX.

Die auf den Servern vorgehaltenen Dokumente sind ganz unterschiedlicher Provenienz: Neben in der Regel

allel dazu wird im Projekt die Frage untersucht, wie eine zukünftige Meta-Informationsebene für elektronische Dokumente sinnvoll zu gestalten ist und inwieweit die dabei vorzusehenden identifikations-, retrieval- und zugriffsrelevanten Informationen maschinell auf Basis der elektronischen Dokumente selbst generierbar sind. Bei dieser Neugestaltung der Erschließungsebene werden Ergebnisse derzeit international geführter Diskussionen berücksichtigt, so etwa die Diskussion um das »Dublin Core Set« und die Arbeit der Text Encoding Initiative (TEI).

Auch das Verhältnis strukturierter Erschließungstechniken im Sinne von Metainformationen zu konkurrierenden technischen Ansätzen im Web auf Basis von Indexierungs- und Suchmaschinen und Techniken des relevance-ranking wird in diesem Rahmen zu bestimmen sein.

Einer der Grundsätze von Webdoc war von vornherein, auf Seiten des Endnutzers ausschließlich allgemein verfügbare Zugriffswerkzeuge wie Netscape oder Mosaic und insbesondere keinerlei Pica-spezifische Clientsoftware vorauszusetzen. Alle Retrievalzugriffe auf Webcat werden daher durch ein HTTP-Serversystem, Web-OPC, vermittelt, das damit eine wesentliche technische Grundlage des Projekts darstellt.

Die zweite projektbezogene technische Entwicklung ist bestimmt durch Überlegungen hinsichtlich der Zugriffsbedingungen für elektronische Dokumente und der hiervon abhängigen Berechnungsverfahren. Die auf den lokalen Servern aufliegenden Dokumente unterliegen in der Praxis unterschiedlichen Zugriffsbedingungen: So finden sich dort Dokumente in der akademischen public domain (diese sind frei zugänglich) neben kommerziellen Verlagsprodukten. Um dieser Situation gerecht zu werden, sieht der Projektplan drei Kategorien von Dokumenten vor:

1. Dokumente, die nicht dem Copyright unterliegen und die ohne Bezahlung verfügbar gemacht werden. Die in den Dokumentbeschreibungen des Webcat enthaltene URL kann in solchen Fällen ohne weitere Zwischenschritte und Vermittlungsinstanzen als Hyperlink auf das gewünschte Dokument aktiviert werden. Dennoch kann es auch hier erforderlich sein, Statistikinformationen zu generieren.

2. Copyrightpflichtige Dokumente, für deren Benutzung bezogen auf eine

# Chaos im Web

:// das Chaos unklarer und zum Teil auch miteinander unvereinbarer kommerzieller Zugriffsbedingungen.

Webdoc ist der Versuch, angesichts dieser Problemfelder traditionelle Stärken der Bibliotheken ins Spiel zu bringen und eine von Bibliotheken kontrollierbare Informationsdomäne als Bestandteil des WWW zu etablieren. Diese virtuelle elektronische Bibliothek – nichts anderes ist Webdoc im Grunde – kombiniert Stärken des genuinen Web-Ansatzes mit Strukturdiensten für die Informationsverwaltung in einem transparenten und flexiblen Dienstkonzept. Im Kern handelt es sich um ein Projekt zur Qualitätssicherung im WWW.

## Webdoc: Technik und Service

Webdoc ist ein zu Beginn des Jahres 1995 von der Pica-Stiftung in Leiden initiiertes Projekt, an dem Bibliotheken und Informationsanbieter in den Niederlanden, in Deutschland und in den USA beteiligt sind.

Der Kernansatz von Webdoc ist ein Dienstkonzept, es handelt sich weniger um ein technisch orientiertes Projekt als vielmehr um den Versuch, im technischen Kontext des Internet und des WWW eine neuartige Dienstleistung aufzubauen und dabei traditionelle Stärken der Bibliotheken als Informationsanbieter zu aktivieren.

kostenfrei nutzbaren universitären Publikationen wie Dissertationen und ähnlichem finden sich dort auch copyrightpflichtige Verlagsprodukte (derzeit typischerweise elektronische Zeitschriften). Die Inhalte dieser lokalen Dokumentenserver werden ganz im Sinne konventioneller bibliografischer Beschreibungen in einer gemeinsamen Nachweisdatenbank, Webcat, erschlossen.

Für den Nachweis in Webcat ist ein von Pica und der Königlichen Bibliothek in Den Haag gemeinsam entwickeltes Format, das sogenannte E-Format, maßgeblich. Dieses Format basiert auf dem allgemein bei Pica-Anwendern üblichen bibliografischen Erfassungsformat und definiert zusätzlich eine Reihe von Modifikationen und Zusätzen für die Erschließung elektronischer Dokumente. Insbesondere enthalten die Nachweisdaten in Webcat sozusagen als Signaturäquivalent die Hyperlinks zu den lokal vorgehaltenen Dokumenten, die URLs.

Das E-Format sollte nicht als der Versuch mißverstanden werden, den neuartigen Erschließungsanforderungen elektronischer Dokumente in vollem Umfang gerecht zu werden. Es handelt sich vielmehr um einen pragmatischen Definitionsrahmen, der den kurzfristigen Aufbau der Webcat-Datenbank möglich machen soll. Par-



## BIBLIOTHEKEN

definierte Nutzergruppe eine Lizenz-Rahmenvereinbarung abgeschlossen wurde. In diesen Fällen sind nutzergruppenbezogene Verfahren für die Zugangsadministration und darauf bezogene Statistikinformationen erforderlich.

**3. Copyrightpflichtige Dokumente mit auf den Einzelzugriff bezogener Individualberechnung.** Für solche Fälle ist eine Zahlungsfunktion mit endnutzerspezifischer Kontenführung vorgesehen. Auch hier werden Statistikinformationen erforderlich.

Zusätzlich können solche Zugangsbedingungen global für alle Dokumente eines Anbieters, für eine Dokumentklasse spezifisch oder individuell dokumentbezogen definiert sein. Diese Definitionen können, bezogen auf Nutzergruppen oder Einzelnutzer, variieren. Und schließlich spielen in diesem Umfeld sicherheitsrelevante Überlegungen eine besondere Rolle.

Diesen Tatsachen trägt der technisch aufwendigste Entwicklungsbestandteil des Projekts Rechnung: der zentrale licensing and accounting server. Dieser Server verwaltet auf die Teilnehmerbibliotheken bezogene Lizenzinformationen. Im Rahmen entsprechender Vereinbarungen für die Lizenzierung können dann Endnutzer der jeweiligen Einrichtungen Zugang zu den Dokumenten erhalten.

Anderenfalls wird eine Direktabrechnung erforderlich. Für diese Zwecke wurde der schon im Rahmen von niederländischen Rapdoc-Projects implementierte Mechanismus für die Verwaltung von Vorauszahlungen (deposit accounts) erweitert. Dabei erhält der Endnutzer für eine hinterlegte Geldsumme eine Nutzerkennung und eine Anzahl von Pincodes. Im Rahmen von Webdoc berechnen sie zum Zugriff auf Dokumente.

Schließlich wurde eine sicherheits- und zahlungsrelevante Erweiterung der WWW-typischen Zugriffsmethode über URL definiert. Dabei finden Verschlüsselungstechniken (RSA) Anwendung.

Der Zugriff auf lizenzpflichtige Dokumente unter Nutzung der Zusatzentwicklungen in Webdoc verläuft damit folgendermaßen:

1. Die URL des gewünschten Dokuments wird dem accounting server übermittelt.
2. Dieser Server bestimmt sowohl die Kosten wie die Zugriffsbedingungen für das Dokument.
3. Das Serversystem überprüft, ob der Endnutzer entweder einer zugriffsbe-

rechtigten Nutzergruppe angehört (in diesen Fällen werden die abrechnungsrelevanten Daten gruppenbezogen kumuliert) oder ob sein deposit account ausreichend hoch ist.

4. Nach positivem Abschluß des letztgenannten Schritts wird die URL des gewünschten Dokuments um eine Reihe von Zusatzinformationen ergänzt (IP-Adresse des Endnutzers, dessen ID, Transferkosten, Berechnungsmethode, eindeutiger Zeitstempel und eine mit einem public-key-Mechanismus generierte digitale Unterschrift); die Kombination dieser Informationen wird als Golden URL (GURL) bezeichnet, gespeichert und dem Dokumentanbieter übermittelt.

5. Der Dokumentanbieter hat als Ergänzung seines Dokumentservers eine von Pica entwickelte Clientsoftware implementiert, die die formale und inhaltliche Konsistenz der empfangenen GURL überprüft.

6. Nach positivem Abschluß dieser Überprüfung erhält der Endnutzer Zugang zum elektronischen Dokument, parallel wird die GURL als Bestätigung des erfolgten Zugriffs an den accounting server zurückgesandt, der die Transaktion dann gegebenenfalls mit einer Belastung des deposit account abschließen kann.

Dieser komplexe Mechanismus operiert fast ausschließlich im Hintergrund. Für den Endnutzer ist der Unterschied zu einem Internet-üblichen Dokumentzugriff minimal: Einzig Abbuchungsoperationen vom Konto des Endnutzers setzen (natürlich) dessen Bestätigung und damit ein Mehr an Interaktion voraus.

Um eines der zentralen Charakteristika dieses Dienstes, die stabile Verfügbarkeit der Dokumentressourcen, verlässlich zu sichern, werden in Webdoc zusätzlich Überlegungen angestellt, die Verfügbarkeit der lokalen Serverressourcen durch einen zyklischen Hintergrundprozeß laufend zu überprüfen und gegebenenfalls temporär nicht aktivierbare Links in Webcat stillzulegen.

Im Kern also ist Webdoc ein Service, der mittels zentraler Erschließung und zentral implementierter Lizenzierungsmechanismen den Endnutzerzugriff auf verteilte elektronische Dokumentressourcen erlaubt.

Webdoc ist damit eine Art stabiler und strukturierter Sektor des Internet, der durch transparente und variabel gestaltbare Konditionen, klare Lokalisierung und eine Metaebene der Erschließung definiert ist. Die-

se Charakteristika kompensieren einige der grundlegenden Negativaspekte des Web-Ansatzes und reaktivieren zugleich klassische bibliothekarische Qualitäten.

### Webdoc: Projektorganisation und Projektpartner

Webdoc ist ein Projekt der Pica-Stiftung in Leiden. Die folgenden Einrichtungen sind am Projekt beteiligt.

#### In den Niederlanden:

- Die Universitäten von Amsterdam, Groningen, Nijmegen, Limburg, Tilburg, Utrecht, Twente, Rotterdam und Brabant
- Die Königliche Bibliothek in Den Haag
- Kluwer Academic Publishers
- Pica, Zentrum für Bibliotheksautomation

#### In Deutschland:

- Die Universitäten von Göttingen, Hamburg, Hannover, Kassel und Oldenburg
- Die Deutsche Bibliothek
- GBV – Gemeinsamer Bibliotheksverband der Länder Bremen, Hamburg, Mecklenburg-Vorpommern, Niedersachsen, Sachsen-Anhalt, Schleswig-Holstein und Thüringen.

#### In den USA:

- Mehrere in der Research Libraries Group kooperierende Bibliotheken
- Academic Press.

Die hier genannten Institutionen bilden den Rumpf des Webdoc-Projekts. Sie entscheiden gemeinsam über strategische Fragen der Projektdurchführung in einem General Steering Committee, dem die Leiter der Einrichtungen angehören.

Wir sind zu Projektbeginn von der Annahme ausgegangen, daß auf dem Markt konkurrierende Verlage wohl kaum im Rahmen einer absolut transparenten Projektorganisation kooperieren würden – sie müßten dann nämlich bereit sein, ihre Statistikdaten und zumindest Teile der Internalkulation den Konkurrenten offenzulegen. Daher wurde die Kooperation mit Verlagen in Unterprojekten angesiedelt, die in Kooperation mit wichtigen wissenschaftlichen Verlagen entweder bereits laufen oder in Verhandlung sind.

Im Rahmen einer globalen technischen Kooperation arbeiten Pica und RLG, die Research Libraries Group in den USA, beim Aufbau eines integrierten Systems für Informationsretrieval und -Lieferung zusammen. In



diesem Rahmen entsteht in den USA ein funktional-technisch eng verwandter und mit Webdoc kompatibler Dienst. Der Austausch von Bestandsnachweisen zwischen diesen beiden Diensten ist Bestandteil der Kooperationsvereinbarung zwischen RLG und Pica. Daneben ist RLG seit Oktober 1995 aktiver Teilnehmer des Webdoc-Projekts; eine Untergruppe der RLG-Bibliotheken ist direkt an der Pilotphase des Projektes beteiligt.

Das Projekt wurde 1995 begonnen und ist auf eine Laufzeit von zirka 30 Monaten angelegt, wird also dieses Jahr abgeschlossen sein. Die ersten beiden Projektphasen (Technische Entwicklung und Dienstintegration) sind abgeschlossen, auch die lokale Infrastruktur (Dokumentserver, Netzeinbindung) steht inzwischen in allen Bibliotheken zur Verfügung.

Derzeitige Hauptaktivitäten sind der Aufbau der lokalen Serverinhalte, die Dokumentakquisition (akademische, Public-domain- und Verlagsprodukte) und die Erschließung dieser Inhalte in der Webcat-Datenbank. Die zunächst noch beschränkte Freigabe des Webdoc-Dienstes für die Pilotphase ist im Mai erfolgt. Eine frei-

zugängliche Demonstrationsumgebung steht unter <http://www.pica.nl> zur Verfügung.

### Webdoc: eine Chance für Bibliotheken und Verlage

Bücher und Bibliotheken werden wohl kaum so bald obsolet oder museal sein, wie dies von manchen derzeit gern suggeriert wird. Es geht wohl auch kaum an, WWW-basierte Nachweisdienste wie Yahoo! ohne weiteres mit einer virtuellen elektronischen Bibliothek gleichzusetzen.

Allerdings werden Bibliotheken aktiv (und nicht rein defensiv) den im Gang befindlichen Wandel in der Publikations- und Distributionspraxis mitgestalten müssen, wenn sie sich nicht in bestimmten Wissenschaftssegmenten bald tatsächlich ausgehebelt in einer musealen Nische wiederfinden wollen. Die Beteiligung an Projekten wie Webdoc ist in diesem Sinne ein wesentlicher Beitrag für die Neudefinition der Rolle von Bibliotheken als qualitätssichernde und stabilisierende Instanzen bei der Gestaltung von Diensten zur Vermittlung elektronischer Publikationen.

Webdoc ist jedoch zugleich eine Chance für die Neugestaltung der Zu-

sammenarbeit von Verlagen und Bibliotheken im Zusammenhang der rapide gewandelten elektronischen Publikations- und Rezeptionskette. Denn ein Gutteil der von Bibliotheken in ihrer Vermittlerrolle beklagten Defizite WWW-basierter Informationsdienste stellt auch Verleger vor massive Probleme. So zum einen die mangelhaften Mechanismen für die Authentifizierung und Statussicherung elektronischer Dokumente. Die unkontrollierte Dynamik elektronischer Dokumente im Verein mit den mehr als ungenügenden Mechanismen für deren Qualitätssicherung stellt vielleicht eine der derzeit aktuellsten Herausforderungen für wissenschaftliche Verlage dar. Gängige WWW-Suchwerkzeuge indexieren weitgehend undifferenziert und sind zu einer qualitativen Bewertung der indexierten Referenzen kaum imstande. Das Ergebnis sind aus Nutzersicht Treffermengen, in denen hochwertige, professionellen Maßstäben genügende Publikationen gleichwertig neben dem Inhalt elektronischer Papierkörbe präsentiert werden. Hier könnte Verlagen eine neue Rolle zuwachsen, abgeleitet aus ihrer klassischen Funktion als Instanzen der Selektion und der Qualitätssicherung. Die zukünftige Bedeutung des Publikationsbegriffs wird im Umfeld von Projekten wie eben Webdoc mitgestaltet. Eine Zusammenarbeit von Verlagen und Bibliotheken ist dafür unverzichtbar.

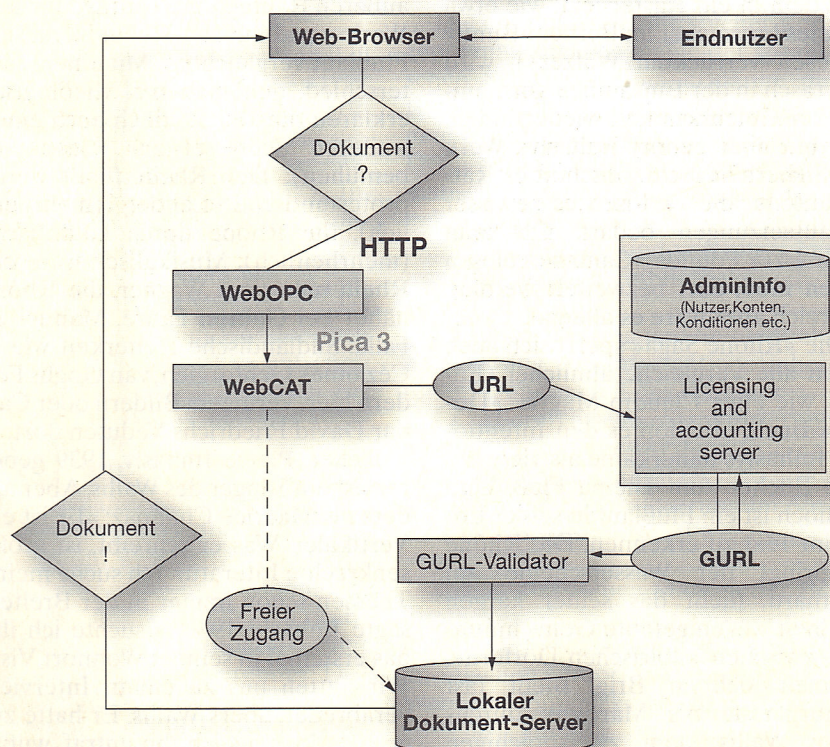
Zum anderen dürfte das Chaos der kommerziellen Zugriffsbedingungen im Umfeld WWW-basierter Dienste Verlagen Probleme bereiten, im Vergleich zu denen die oben angesprochenen bibliothekarischen Aspekte banal wirken. Webdoc bietet einen flexiblen, in das WWW-Umfeld integrierten Mechanismus für die Verwaltung und Vermittlung von Lizenzinformationen mit integrierten Zahlungsfunktionen. Dieser Aspekt kann von Verlagen für die Neugestaltung von Distributionswegen unter den Bedingungen des elektronischen Publizierens nutzbar gemacht werden.

So wie die klassische, konventionell arbeitende Bibliothek ohne verlegerische Partner völlig undenkbar war (und ist), werden auch künftig entstehende elektronische Bibliotheken ohne leistungsfähige Verlage als Partner nicht arbeitsfähig sein. Deshalb ist Webdoc programmatisch als Projekt für die Kooperation von Bibliotheken und Verlagen angelegt und bietet beiden die Chance zur Neudefinition einer leistungsfähigen Allianz.

## WebDOC: Funktionsdiagramm

Quelle: GBV

Zusammenspiel von generischen  
und Pica-eigenen Komponenten



Grafik: Ulrich Schmidt-Bensheim